Autumn Conference 2017
Paper 6

Authors:
Simon Withey, Environment Agency
Keith Davis, Environment Agency
Matt Wheeldon, Wessex Water

21st Century Drainage Programme – Storm Overflows Environmental Impact Assessment

Abstract

The need to monitor the performance of storm overflows and develop a robust process for managing high frequency spillers was set out by Government in 2013. In response to this challenge, the Environment Agency has worked with the water industry under the 21st Century Drainage Programme to develop a framework for assessing the impact of high frequency spillers, and for evaluating the costs and benefits of improvement measures, including wider socio-economic considerations. The aim of this Storm Overflows Assessment Framework (SOAF) is to ensure the industry has a process in place which can demonstrate that storm overflows are being monitored, and managed, in light of the pressures of growth, urban creep and changing rainfall patterns. As public information on overflow operation becomes more widely available, the SOAF will demonstrate to customers and non – governmental organisations (NGOs) that the industry is proactively investigating the impact of its storm overflows, and taking action where necessary. The SOAF is also intended to show that the United Kingdom has a process in place which can demonstrate that sewerage systems are compliant with relevant legislation such as the Urban Waste Water Treatment Regulations.

This paper describes the investigational stages of the SOAF. These include the process of reviewing event duration monitoring data to identify high frequency spillers for investigation, and how environmental impact is quantified and scored to aid the later assessments of costs and benefits. The environmental impact assessment involves three components – aesthetic, biological (macro invertebrate), and water quality impact. Each of these components is scored and classified separately depending on the information available, and will link to the cost and benefits framework. For the biological and water quality impacts, a hierarchical approach is adopted which gives preference to biology data over modelled water quality assessments. Where water quality modelling is involved, traditional UPM approaches are adopted with the fundamental and 99 percentile standards used as a reference to quantify impact.

1.0 Introduction

Discharges from storm overflows have become a reputational issue for the water industry. Over the last decade public awareness of the presence and impacts of combined sewer overflows (CSOs) has increased following several national media campaigns by organisations such as the Marine Conservation Society (MCS), and Surfers Against Sewage (SAS). In September 2009 the BBCs investigative program Panorama examined the impact of sewer overflows during the episode “Britain’s Dirty Beaches”. Although media interest was mainly concerned with impacts at sensitive locations such as bathing and shellfish waters, wider concerns were raised about the performance of CSOs, and how companies manage their assets. Information requests by MCS at the time highlighted that only 25% of CSOs were monitored, and that much of this data was not available to the public or regulators. As a result, the MCS called for better public information on overflow operation to be provided through the expansion and standardisation of CSO spill monitoring and reporting (MCS, 2011).

At a similar time to the Panorama program, the European Commission (EC) announced that it was taking the United Kingdom (UK) to the European Court of Justice (ECJ) on the grounds that sewer networks and treatment plants serving London (Beckton & Crossness), and Sunderland (Whitburn), did not comply with the requirements of the Urban Waste Water Treatment Directive (UWWTD). The Directive requires sewer networks for agglomerations with a population equivalent of 2,000 or more to be designed, constructed and maintained according to best technical knowledge not entailing excessive cost (BTKNEEC). This includes the volume and characteristics of the wastewater, the prevention of leaks, and the limitation of pollution of receiving waters due to storm overflows. Although the Directive acknowledges the need for CSOs, they should only operate during "unusually heavy rain". The EC argued that the discharges of storm sewage were too frequent and in excessive quantities. In October 2012 the ECJ agreed that the UK had failed to provide adequate sewerage systems for London and Sunderland (EUR-Lex, 2012). Storm discharges were far too frequent to be exceptional in nature, and the UK had not demonstrated that the cost of reducing discharges was disproportionate to the benefits a more effective drainage system might provide. In May 2017 the UK was again judged to have provided inadequate collection systems, this time for Gowerton and Lanelli in Wales (EUR-Lex, 2017).

As well as the current reputational and legal issues, pressures on CSOs are only expected to increase. Population growth, urban creep and changing rainfall patterns all risk increasing the frequency and volume of storm discharges (UKCCRA, 2017; Water UK, 2017).

In light of the ECJ ruling, wider public awareness, and the long-term pressures on CSOs, Government wrote to the water companies in July 2013 setting out its expectations for CSO management (Benyon, 2013). Firstly, Government requested the expansion of monitoring so that the vast majority of CSOs are monitored by 2020. Secondly, measures were requested to address high frequency discharges identified through monitoring, or where discharges are otherwise unsatisfactory. Each company was asked to set its own objectives and trajectory for dealing with high frequency CSO discharges, responding to existing data as well as that derived from the expansion of monitoring to 2020.

Given the importance of a standardised and transparent monitoring program, the expansion of event duration monitoring (EDM) was included within the permitting regime, and supported through the National Environment Programme of the 2014 Periodic Review. A risk based approach was developed to deliver this, with storm overflows categorised according to discharge significance, based on amenity, spill frequency and the provision of information (Environment Agency, 2013). Discharges identified as either high or medium significance under the criteria, which includes the vast majority of storm overflows, require EDM. Monitors will record the frequency and duration of storm discharge events, and will be installed at CSOs on the network, at sewage treatment works inlets, and on storm tanks.

In July 2015 a group was set up under workstream 3 of the 21st Century Drainage Programme (Water UK, 2017). The group included representatives from the environmental regulators and water and sewerage companies in England and Wales, and was tasked with working together to develop a process for addressing high frequency discharges. Over the last two years the group has developed a draft 5 stage process to achieve this known as the Storm Overflows Assessment Framework (SOAF). A flow chart showing the main stages involved is shown in Figure 1. High frequency discharges are identified for investigation during Stage 1 of the process. An assessment of environmental impact is made under Stage 2, while the costs and benefits of reducing spill frequencies are quantified under Stage 3. The evaluation of benefits considers both environmental impact and wider socio – economic considerations. Investment decisions using this BTKNEEC assessment will be determined and implemented under Stages 4 and 5.

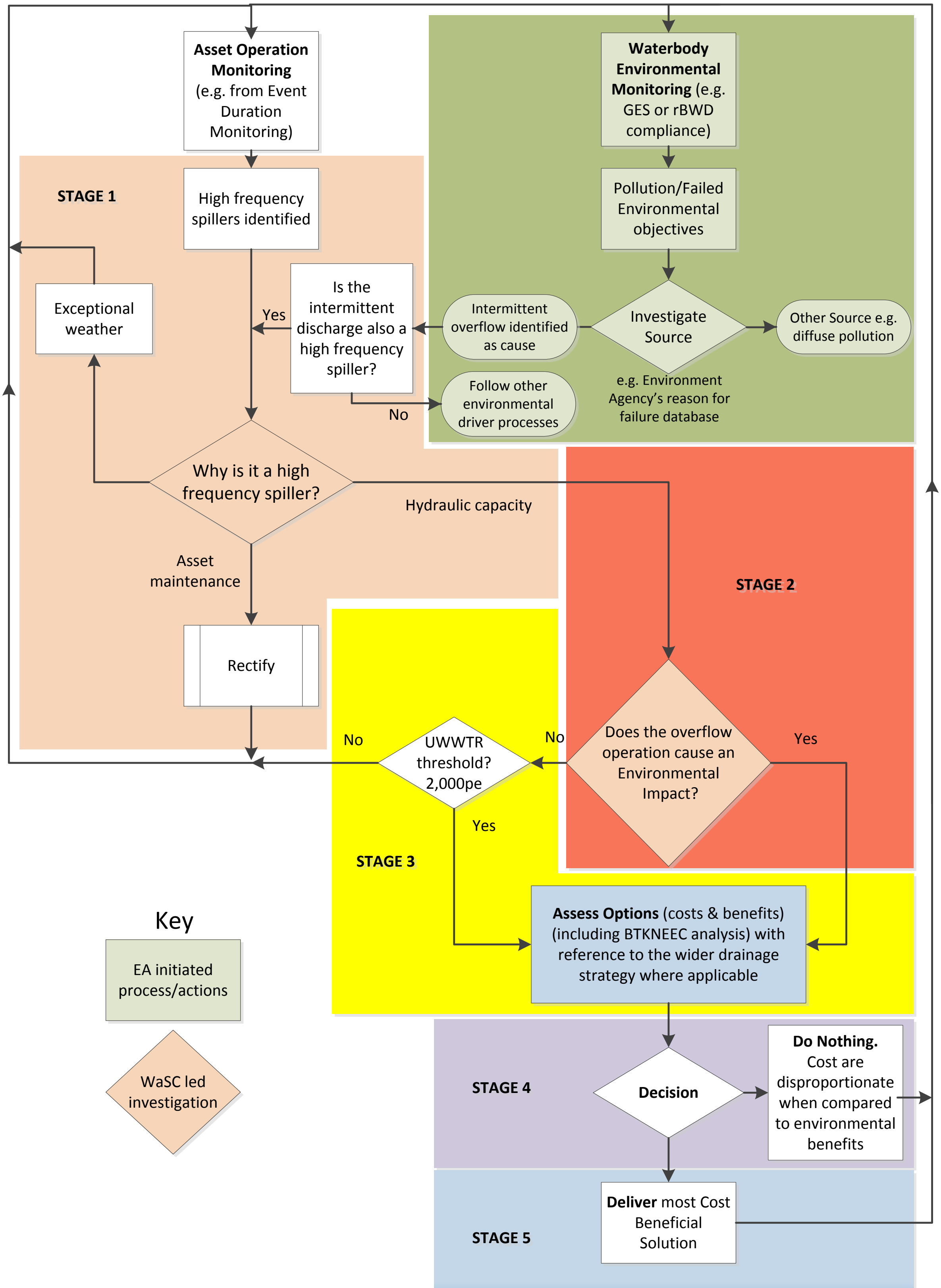

Figure 1. Assessment framework for addressing high frequency discharges from storm overflows under the Urban Waste Water Treatment Regulations (UWWTR).

Through the BTKNEEC assessment, the SOAF should allow the UK to be able to demonstrate that the requirements of the UWWTD are being met. As the United Kingdom leaves the European Union (EU) the legal and legislative landscape may change. However, the purpose of the SOAF is wider than compliance with current legislation. It will allow the industry to demonstrate that it has a process in place which shows that storm overflows are being monitored, and managed, in light of the pressures of growth, urban creep and climate change. As public information on overflow spill frequencies becomes more widely available, the SOAF will also demonstrate to customers and non – governmental organisations (NGOs) that the industry is proactively investigating the impact of its storm overflows, and taking action where necessary.

This paper describes stages 1 and 2 of the SOAF, which involve the identification of high frequency spillers, and quantification of their environmental impact. The subsequent assessment of cost and benefits under stage 3 is described separately in Paper 4 of this conference.

2.0 Stage 1 – Why is the storm overflow a high frequency spiller?

Once annual EDM reports become available, storm overflows will be identified for investigation using the spill frequency triggers in Table 1, which vary according to the number of years' of EDM data available.

Table 1. Spill frequency investigation triggers.

No. of years EDM data	Investigation trigger (average no. of spills / year)
1	>60
2	>50
3 or more	>40

The triggers were chosen based on predicted spill frequency data for overflows from recent (AMP5) UPM studies in some Northwest and Yorkshire catchments, where rainfall is higher than the national average. Based on this data it was thought that roughly 10 – 40% of overflows in these wetter catchments might exceed the triggers, depending on the trigger and the catchment's annual average rainfall (which varied from roughly 900mm – 1200mm). These triggers sought to begin the process of addressing 'the worst offenders' without creating an unmanageable number of investigations, particularly in the first year. It was acknowledged that the triggers would disproportionately affect companies on the western side of the UK, which experience higher annual rainfall. However, although percentile cuts were considered, it was felt that absolute spill frequencies are important to the public and other stakeholders, and easier to communicate.

Based on existing EDM data, companies estimated that perhaps 7 – 10% of overflows will trigger investigation. However, estimates vary between companies and are affected by a range of factors including the annual rainfall in their region, and whether existing EDM is biased toward coastal protected areas where many overflows have already been improved. With inland EDM programs yet to be fully implemented (2020), the exact scale remains unknown.

Once a storm overflow has been identified as a high frequency spiller through EDM, the first stage in the SOAF involves working out why the spill frequency trigger was exceeded. Firstly, catchment rainfall is reviewed to determine whether rainfall was exceptional during any of the EDM reporting years. If rainfall was exceptional, the relevant calendar year of EDM data is flagged accordingly and not used. The remaining years of EDM data are then reassessed against the triggers (Table 1). There are two options for assessing whether rainfall was exceptional:

A) Environment Agency water situation reports

Water situation reports are published each month on gov.uk at a hydrological area level and include rainfall statistics for the last 12 months. Rainfall is typically reported as a percentage of a long-term annual average figure and categorised according to 7 classes, from 'exceptionally low' through to 'exceptionally high'. The probability ranking used to define the categories for rainfall is updated every five years as the period of long-term observed data increases. If rainfall during the year was classified as 'exceptionally high', then that year's EDM data is not used in assessing whether the trigger has been exceeded. The 'exceptionally high' category is defined as a value that is likely to fall within the band for 5% of the time.

B) Local rainfall records

Water situation reports provide an indication of exceptional rainfall at a hydrological area level. They may not be representative of some local catchments within that area. Consequently, an alternative is to use local rainfall data for the catchment where available and use this to estimate if rainfall during the EDM calendar year was exceptional (5% probability). This may be carried out to varying levels of complexity, but at an overall annual resolution. For example, the total depth of rainfall recorded by the gauge in the year could be compared to the long-term record to decide whether it was exceptional. Alternatively, the number of rainfall events of the same critical duration of the overflow could be counted and compared to the long-term annual average for those events. However, individual spill events or time periods within the annual dataset should not be excluded.

As well as checking whether the catchment had been exceptionally wet, companies will also inspect the asset and potentially parts of the upstream and downstream catchment to see whether the high spill frequency is the result of asset condition or maintenance issues. Asset condition includes a wide variety of problems that might be contributing to frequent discharges, for example partial blockages, structural defects, worn pumps, infiltration and surface water misconnections. If available, existing verified hydraulic models, survey data and maintenance records will be used to assist with this assessment. Where asset condition is found to be making a significant contribution to spill frequency, the problem is resolved outside of the SOAF. The year, or years of EDM data affected by poor asset condition are archived, and not used for ongoing assessment against the triggers.

By excluding impact investigations where frequent discharges are due to exceptional rainfall or poor asset maintenance, the SOAF will concentrate on those assets which discharge frequently in typical wet weather, simply as a function of their permitted hydraulic design and upstream contributing area.

3.0 Stage 2 – Does the overflow cause an environmental impact?

The second stage of the SOAF attempts to quantify the environmental impact of the overflow, in order to provide data for use in the cost – benefit assessment under Stage 3. The impact assessment involves 3 components, which are assessed and scored separately:

- Aesthetic impact
- Invertebrate (biological) impact
- Water quality impact

The assessment is hierarchical and gives preference to invertebrate impact data over modelled water quality assessments. The process is summarised in Figure 2. It begins with the aesthetics assessment (Stage 2a). A score and classification ranging from 'no impact' to 'severe impact' is assigned to the overflow.

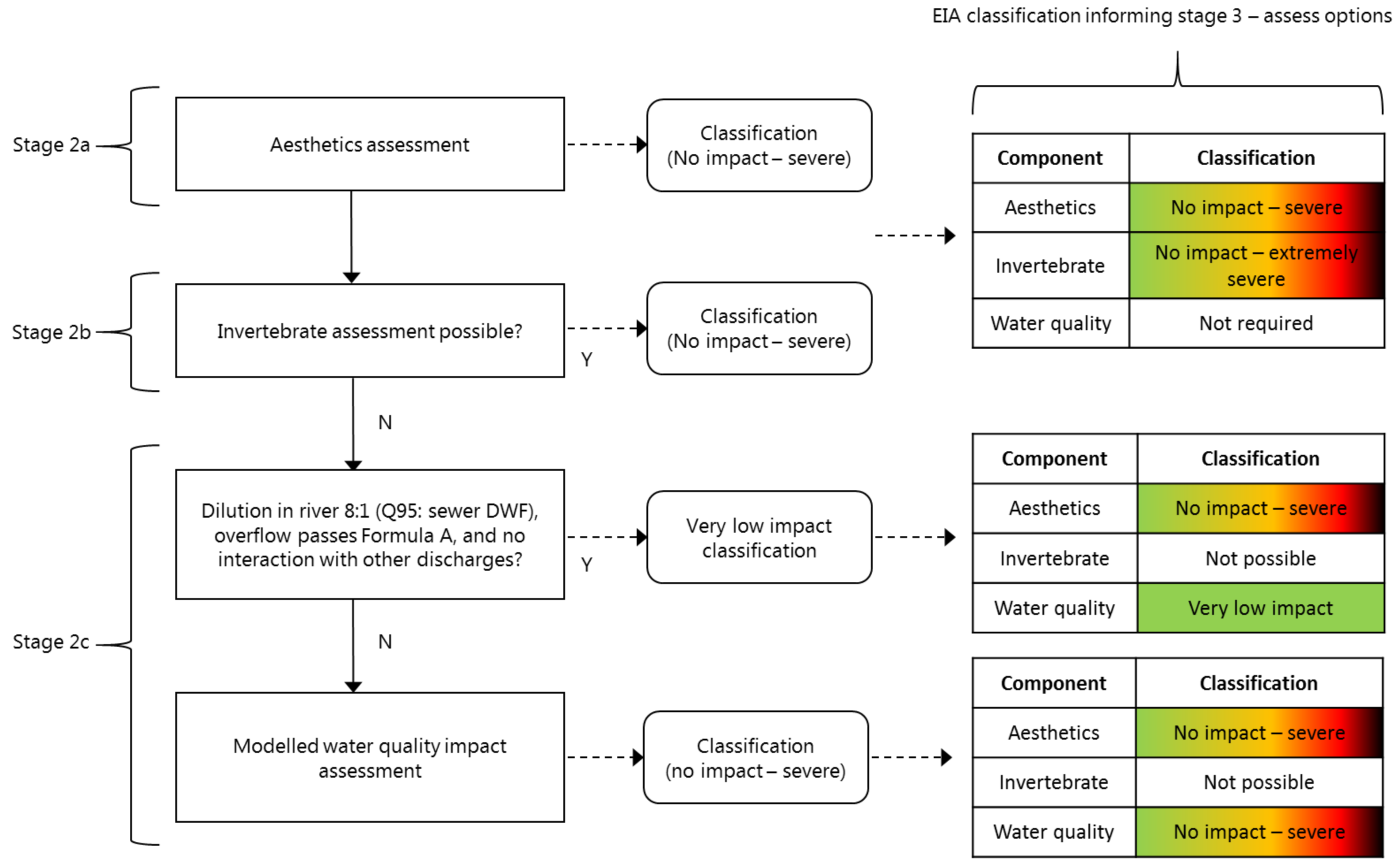

Figure 2. Summary of environmental impact assessment process.

Where it is possible to collect representative invertebrate samples upstream and downstream of the overflow, an invertebrate impact classification is then assigned under Stage 2b, ranging from 'no impact' to 'extremely severe'. If the invertebrate assessment is possible, an assessment of water quality impact is not needed. Although UPM traditionally involves a modelled assessment of impact against water quality design standards intended to protect aquatic ecology, the SOAF promotes invertebrate sampling over modelling for two reasons. Firstly, invertebrates are good indicators of water quality, and have been used to detect the impact of intermittent discharges. It was felt that invertebrate data would provide direct evidence of impact. The disadvantage might be where the critical reach for storm sewage impact is a long distance downstream of the overflow, and this is not detected by invertebrate sampling closer to the outfall. Secondly, invertebrate sampling offered a relatively inexpensive method of quantifying impact, which would avoid or reduce the need for potentially expensive water quality modelling where no impact is found.

For many overflows, it will not be possible to collect representative samples upstream and downstream of the outfall. This will be the case in many urban catchments where multiple outfalls arise in close proximity, or where CSOs overflow to surface water sewers before entering the watercourse, and it is not possible to confidently attribute impacts to the specific overflow under investigation. In these scenarios an assessment of water quality impact is required under Stage 2c. This is likely to be the case for the majority of overflows investigated under the SOAF. The water quality impact assessment involves an initial assessment based on dilution. If the dilution criteria are not met, a modelled impact assessment is required. The overflow is then assigned a water quality impact classification ranging from 'no impact' to 'severe impact'.

More detail on the individual components is described in stages 2a – 2c below.

3.1 Stage 2a – Aesthetics

The purpose of this first stage in the hierarchy is to address issues surrounding the public acceptability of storm overflows in terms of visible impact, the amenity value of the receiving water, any history of substantiated public complaint, and any history of recorded pollution incidents due to storm sewage discharges. This stage involves a visual inspection of the outfall and receiving water consistent with traditional approaches such as the FR0466 survey guide (FWR, 1994). A desktop review of customer complaint and pollution incident data will also be carried out. The impact assessment is shown in Table 2 below. Each subcomponent (aesthetics, customer complaint and pollution incidents) is scored separately and the individual scores summed to give a total score. The impact classification according to this total score is shown in Table 3.

Although aesthetics surveys will be based on the historic FR0466 methodology (FWR, 1994), the scoring system in Table 2 above will be used. To allow for site – specific circumstances some judgement will also be needed. For example, where it is foreseeable that litter may be stranded and visible in areas downstream of the notional 50m survey area defined in FR0466, the survey will be extended to include this area. This will be important where the amenity class increases downstream of the immediate 50m reach. For example, where there is a park alongside the watercourse 300m downstream of the outfall, then this would be included in the aesthetics assessment.

Due to the potential effects of bankside vegetation on access, visibility and the potential for litter to collect, two surveys will be required to judge aesthetic impact. One survey should take place in late autumn, winter or early spring (November – April) when any bankside vegetation is minimal, and one survey should take place between late spring and early autumn (May – October). The worst score returned by the two surveys will be used.

Where more than one overflow impacts the outfall (e.g. where discharges are made to a common surface water sewer), judgement will be required in order to assign a score taking into account whether the overflows are screened and their comparative spill volumes.

Amenity value will be recorded during the survey and will be an important input to the cost – benefit assessment. Amenity will be recorded according to the highest amenity class within 1km downstream of the overflow, or by using judgement as appropriate.

Pollution incidents recorded on the Environment Agency's National Incident Recording System (NIRS) will be used in the assessment (see Table 2). However, as the SOAF is concerned with wet weather performance, only incidents attributed to wet weather discharges from the overflow are included. These incidents are recorded in NIRS with a pollutant type of 'storm sewage'. Pollution incidents in dry weather, for example due to problems such as blockages, will not be included in the assessment.

Table 2. Aesthetics impact assessment.

Aesthetics		**Score**
Sewage derived litter (no. of items) downstream	0	0
	1 – 10	5
	11 – 25	10
	26 – 50	15
	>50	20
Sewage fungus on outfall (present / absent)	N	0
	Y	5
Sewage fungus in downstream mixing zone (% cover)	0	0
	<2	5
	2 – 10	10
	11 – 25	15
	26 – 50	20
	>50	25
Public complaints		**Score**
No. of validated public complaints related to wet weather discharges from the overflow	0	0
	1 – 4	10
	5 – 9	20
	10 – 14	30
	15 or more	40
Pollution incidents due to storm sewage	**Incident category**	**Score (per incident)**
NIRS incidents due to storm sewage attributed to the overflow	Cat 3	20
	Cat 2	40
	Cat 1	60
Total score		?

Table 3. Aesthetic impact classification.

Total Aesthetic Score	Aesthetic impact classification
0	No impact
1 – 5	Very low
6 – 10	Low
11 – 20	Moderate
21 – 40	High
>40	Severe

3.2 Stage 2b – Invertebrate assessment

Following the aesthetics assessment (Stage 2a) above, the next step is to attempt a biological assessment of water quality impacts through invertebrate sampling. Macroinvertebrates are useful indicators of river quality. They are diverse and widespread in freshwater habitats, and relatively cheap and easy to sample and identify. Different species show different sensitivities to environmental stresses such as organic pollution, which allow the type of stress or pollutant to be identified. Most species are sedentary, so their presence or absence at a location within a river can help determine the source of pollution. Their life cycles are also sufficiently long that episodic stresses which have occurred over time can be integrated, including the impact of intermittent discharges (Extence & Ferguson, 1989). Single samples can be sufficient to reveal environmental stresses that have occurred over the previous few months (Clarke et al, 2003). As a result, various indices have been developed over the years which use macroinvertebrates to detect the most common form of pollution – organic pollution, and to classify the ecological quality of rivers (Woodiwiss, 1964; Chandler, 1970; Armitage et al, 1983; Extence & Ferguson, 1989; Whalley & Hawkes, 1997; Wright et al, 2000).

The most recent invertebrate quality indices used for Water Framework Directive (WFD) classification since the second cycle of the river basin management plans, are the abundance weighted Whalley Hawkes Paisley Trigg (WHPT) indices (Whalley & Hawkes, 1997; Paisley et al., 2007; UKTAG, 2014). The method is designed to detect impacts due to organic pollution and is also sensitive to toxic pollutants. It is used to classify rivers in relation to general degradation.

Table 4. Example of abundance weighted WHPT scores for invertebrate families.

Heptageniidae (mayfly) – sensitive to organic pollution and high scoring

Asellidae (water slater) – less sensitive to organic pollution and lower scoring

Taxon	**Abundance categories**			
	1 – 9 (AB1)	**10 – 99 (AB2)**	**100 – 999 (AB3)**	**1000+ (AB4)**
Asellidae	4.0	2.3	0.8	– 1.6
Heptageniidae	8.5	10.3	11.1	11.1

To carry out the WFD assessment, a minimum of two invertebrate samples are collected – one in spring (March – May) and one in autumn (September – November). The invertebrates in each sample are identified to WHPT family level. Each family found is given a score based on its sensitivity to organic pollution and its abundance in the sample. An example is shown in Table 4 above. Some invertebrates such as the Heptageniidae (a family of mayflies) are sensitive to organic pollution and receive a high score. The presence of these groups indicates good water quality. Other groups, such as the Asellidae (water slaters) are more tolerant of organic pollution and receive a lower score. High abundances of these groups indicate poorer water quality.

The WFD assessment involves a 'worst of' approach using two indices. The first index is the number of WHPT scoring families or taxa found in the sample (NTAXA). The second is the average sensitivity score per taxa (ASPT) for the sample. In order to estimate the quality of a river, the method estimates how much the observed quality departs from what might be expected at the river site under 'natural', 'unpolluted' or 'reference' conditions. To do this an environmental quality ratio (EQR) is calculated for each index (NTAXA and ASPT), by dividing the observed index value by the value of the index for the unpolluted reference condition:

$$\text{EQR} = \frac{\text{Value of observed index}}{\text{Value of index expected at undisturbed reference condition}}$$

Index values for the undisturbed reference sites represent the best observed quality available for a range of river types. To obtain these expected reference values of WHPT NTAXA and ASPT for a site, data on site environmental variables is used to predict the invertebrate groups that would be expected to be present if the site is unpolluted using statistical models. The nine environmental variables used to predict the expected invertebrate fauna are shown in Table 5 below (Clarke et al, 2003; EU STAR, 2004).

Table 5. Environmental variables used to predict reference values of WHPT NTAXA and ASPT.

Variable	**Units / description**
National Grid Reference (NGR)	6 figure (100m) NGR (used to predict mean air temperature and range)
Altitude	Metres above mean sea level to nearest 5m
Distance from the source	The distance along the watercourse between the site and its furthest source in km to the nearest 0.1km
Slope	m/km
Discharge category	Estimates of the long-term historical average discharge (m^3/s) using 10 categories (1 = <0.31, 2 = 0.31 – 0.62, 3 = 0.62 – 1.25, 4 = 1.25 – 2.50, 5 = 2.50 – 5.0, 6 = 5.0 – 10.0, 7 = 10.0 – 20.0, 8 = 20.0 – 40.0, 9 = 40.0 – 80.0, 10 = >80.0m^3/s)
Substrate composition	% cover of clay/silt, sand, gravel / pebbles, cobbles / boulders – estimated at the time of sampling and averaged across a minimum of 3 seasons (spring, summer and autumn).
Stream width	m – averaged across 3 seasons reflecting predominant conditions
Stream depth	cm – averaged across 3 seasons reflecting predominant conditions
Alkalinity	mg $CaCO_3$ / l – annual average estimated from a minimum of 3 evenly spaced samples. Monthly values preferred.

EQRs close to 1.0 indicate that invertebrate communities are close to their natural state. The EQR ratios for the different WFD invertebrate status classes are shown in Table 6 below.

Table 6. Environmental quality ratios for WFD invertebrate status.

EQR Values		Invertebrate Status Class
WHPT NTAXA	WHPT ASPT	
>=0.8	>=0.97	High
>=0.68	>=0.86	Good
>=0.56	>=0.72	Moderate
>=0.47	>=0.53	Poor
<0.47	<0.53	Bad

The process of status classification is shown in Figure 3 (UKTAG, 2014). Prediction and classification of invertebrate quality is carried out for each of the individual spring and autumn samples. A mean EQR is then calculated for the two seasons. Overall classification (MINTA) is based on the worst status class assigned for the multi – season mean WHPT NTAXA and WHPT ASPT. In order to assign a level of confidence in the WFD status class calculated, statistical simulation is used to try and quantify uncertainty. There is inevitably uncertainty in observed / sampled values for a site due to factors such as sampling variation and lab processing error (bias). Similarly, there can be errors in the expected values used to calculate EQRs, for example due to errors made when measuring the environmental variables used to predict the expected fauna. Estimates of uncertainty derived from historic sampling programs are used in conjunction with Monte Carlo simulation to build up a distribution of possible observed and expected index values, and hence EQRs. Ten thousand simulation 'shots' are used. An example of a classification report is shown in Figure 4, which shows the percentage number of simulations in each quality class for NTAXA and ASPT, and therefore the most probable status class for the site.

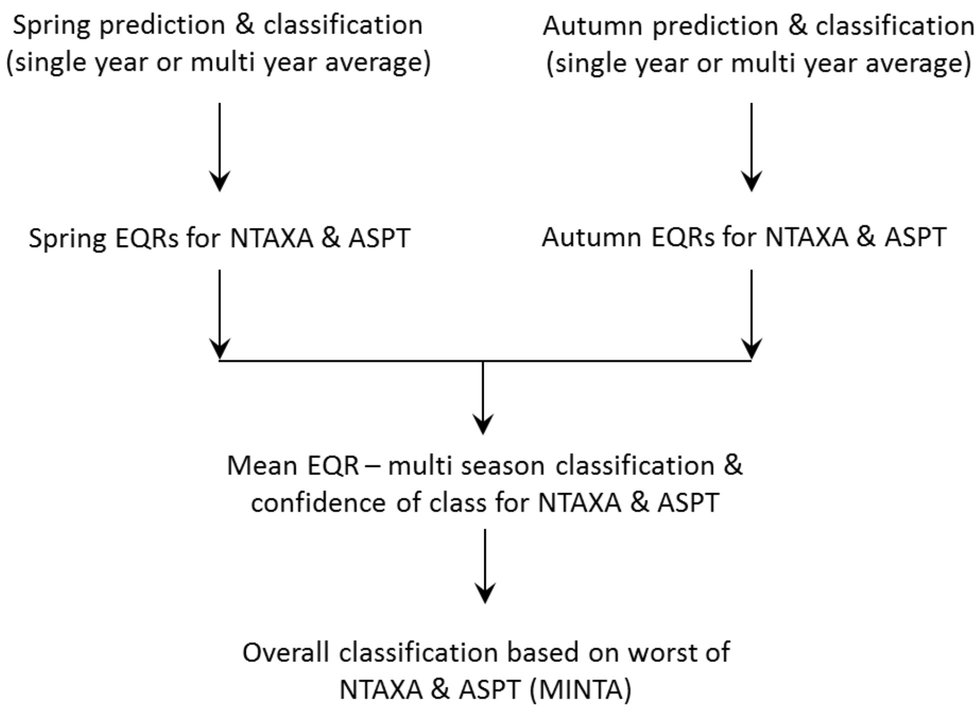

Figure 3. Invertebrate WFD classification process – 'worst of' approach for WHPT NTAXA and ASPT.

Most probable class (high) and associated probability

Percentage of simulations in each class (H = High, G = Good, M = Moderate & P = Poor)

Index	Year	Avg Bias Corr Eqi	Eqr Factor	Avg Fvb Eqr	Class	Prob	Prob H	Prob G	Prob M	Prob P
WHPT Abund MINTA	2017	-	1.0000	-	H	58.81	58.81	41.16	.03	0
WHPT ASPT Abund	2017	0.985	0.9921	0.977	H	58.81	58.81	41.16	.03	0
WHPT NTAXA Abund	2017	1.024	1.0049	1.029	H	99.79	99.79	.21	0	0

Figure 4. Example classification results.

The process of prediction and classification is carried out using software known as the River Invertebrate Classification Tool (RICT). The RICT was developed by the three UK environmental agencies to classify the ecological quality of rivers (Davy – Bowker et al, 2007, 2010). It is hosted by the Scottish Environmental Protection Agency (SEPA) and is available online at https://www.sepa.org.uk/environment/water/aquatic-classification/river-invertebrate-classification-tool/.

Targeted macroinvertebrate sampling and analysis has been used widely in the past in support of UPM studies. The most comprehensive program was carried out by Yorkshire Water during AMP3 investigations in 2000 – 2001. Multi – season sampling was carried out upstream and downstream of approximately 886 CSO outfalls across Yorkshire, to identify overflows likely to be causing impacts for water quality modelling.

The SOAF incorporates a similar approach to that used during the Yorkshire AMP3 investigations, except it uses the latest WHPT indices and WFD quality assessment using the RICT described above. Where possible, macroinvertebrate samples will be collected at representative sites upstream and downstream of the outfall. Samples will be collected in spring and autumn along with data on relevant environmental variables collected over at least at least 3 seasons. This data will be used to carry out the standard WFD quality assessment using the RICT for both the upstream and downstream sampling sites. Quality between the sites will then be compared using the RICT Compare Module, which can be used to assess whether there is a real difference in status class between sites, or at the same site over time. The 'at a glance' report from the Compare Module shows the percentage of simulations where one site (e.g. the downstream site) is one or more quality classes worse than the other site (e.g. the upstream site) for both NTAXA and ASPT. An example of the 'at a glance' results report is shown in Figure 5. For the SOAF this allows an impact scoring system to be developed based on the percentage number of simulations that the downstream site is worse than upstream. This impact scoring system is summarised in Tables 7 and 8. For example, where only 5% of the simulations at the downstream site are one class worse than upstream of the outfall, the impact classification is 'very low'. Where, say, 52% of the simulated results for the downstream site are one class worse than upstream, then the most probable class downstream of the outfall will be worse than upstream and impact is therefore classed as 'high'. The assessment is repeated for both WHPT indices – NTAXA and ASPT, and a 'worst of' approach is used to select the impact classification for input to the benefits assessment. The scoring process will also be repeated for each of the individual spring and autumn samples, and the overall mean of the seasons in order to produce a short – term and long – term impact assessment (Table 9).

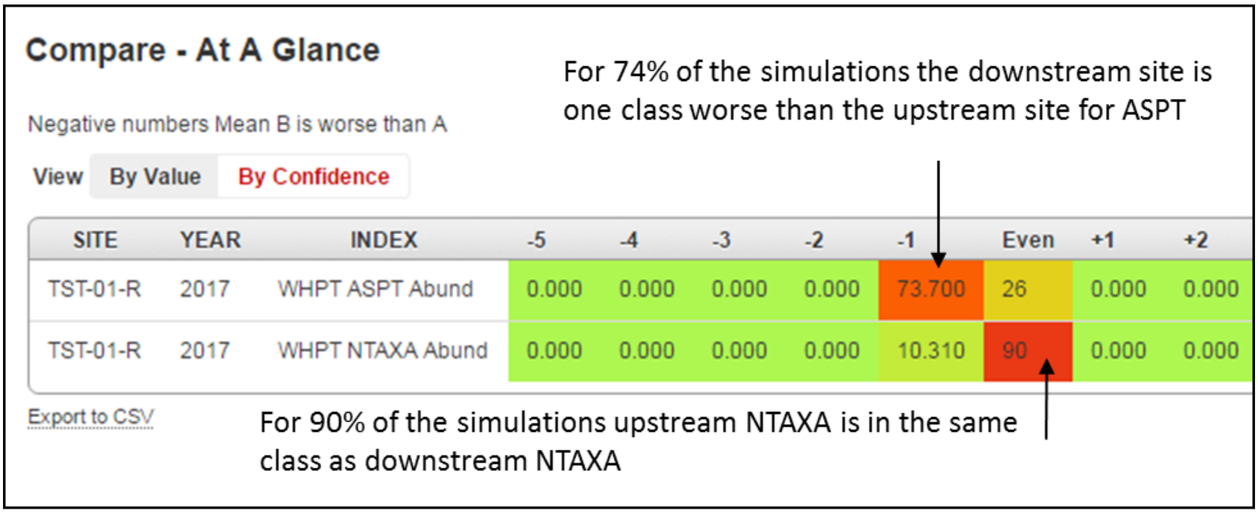

Figure 5. Example RICT compare module 'at a glance' report.

Table 7. Invertebrate impact scoring for WHPT NTAXA and ASPT.

% of simulations the downstream sample is one or more classes worse than upstream	Score	Class Multiplier
1 – 4	1	× No. of classes the downstream sample is worse than upstream
5 – 9	2	
10 – 29	4	
30 – 49	6	
50 – 70	8	
71 – 90	10	
>90	12	

Table 8. Invertebrate impact classification for WHPT NTAXA & ASPT.

Total score	Overall classification
1	No impact
2 – 3	Very low
4 – 5	Low
6 – 7	Moderate
8 – 9	High
10 – 11	Very high
12 – 15	Severe
16 – 19	Very severe
20 or more	Extremely severe

Table 9. Overall short and long – term invertebrate impact assessment.

Type	Description	Value
Short – term	Worst single season classification result for WHPT NTAXA and ASPT	No impact – extremely severe
Long – term	Worst of WHPT NTAXA and ASPT for the overall multi season (spring & autumn) classification	No impact – extremely severe

The invertebrate assessment is only appropriate in simple scenarios where there is a single storm overflow discharging to that reach of the river. Where there are multiple outfalls in close proximity, or other sources of pollution which could account for differences in invertebrate quality between the upstream and downstream sampling sites, then this method will not be used. In degraded urban watercourses where background / upstream invertebrate quality is already poor status then this method will also not be used. As a result, for many storm overflows, water quality modelling will be needed to quantify impact and input to the benefits assessment. This is described in the next section.

3.3 Stage 2c – water quality impact

As discussed above, an assessment of water quality impact is required where it is not possible or appropriate to collect invertebrate samples immediately upstream and downstream of the outfall. The aim of this stage is to quantify the relative impact of a storm overflow on a scale which can be used to input to the cost – benefit assessment.

In order to reduce the need for water quality modelling, an initial screening assessment based on dilution will be used to identify those overflows that are unlikely to be causing water quality issues and jeopardising water quality standards. If the overflow passes forward a retained flow of Formula A over the full duration of spills, the dilution in the receiving water is >8:1 (Q95 river flow: sewer DWF), and there is no potential for interaction with other discharges, then a water quality classification of 'very low' is assigned for input to the benefits assessment.

If the dilution criteria above are not met, then the impact of the overflow on river water quality will be assessed using water quality modelling. The assessment will quantify the impact of the storm overflow on the duration of 99 percentile exceedance, or 99 percentile quality for total ammonia and BOD, and the number of exceedances of the fundamental intermittent standards (FIS) for dissolved oxygen and un-ionised ammonia (FWR, 2012). This should be undertaken as a relative assessment by comparing the impact of the urban drainage system on downstream river quality with and without the discharge from the storm overflow.

It is not expected that complex sewer quality and dynamic river quality modelling is carried out in all cases. The third edition of the Urban Pollution Management (UPM) manual provides guidance on modelling the impact of storm discharges (FWR, 2012). The level of complexity involved depends on the complexity of the problem and the potential cost of any solutions. A complex problem, for example where a large number of storm overflows discharge into a river channel which contains structures such as weirs or sluices likely to affect quality, will need more detailed models and data collection. In contrast, simplified impact approaches will be sufficient for simple scenarios, for example where a single or very small number of overflows discharge into a simple river reach and dilution levels are relatively high. For the purposes of the SOAF, guidance on potential modelling approaches is provided in Table 10. This is intended to assist with scoping individual investigations. There are four levels of complexity:

- Level 1

This is the simplest form of impact assessment. Time series outputs from a verified sewer model are mixed with random picks of upstream river flow and quality selected from statistical distributions. Default or sampled values for storm sewage BOD and total ammonia concentrations can be used and applied as an event mean concentration. The river reach is simplified to a trapezoidal channel. Hydraulic equations are used to estimate the depth and velocity of the mixed flow of river and storm sewage. A simplified water quality model, usually representing the main oxygen demand processes (BOD decay and nitrification) and re-aeration, is used to predict levels of dissolved oxygen and un-ionised ammonia at the end of the reach. Checks against 99 percentile standards and initial un-ionised ammonia can be made at the point of mixing.

- Level 2

This is similar to level 1. However, instead of a stochastic approach to representing upstream river flow, a river flow time series is used. This allows the flow and therefore dilution available in the river at the time of a spill to be better represented. As in level 1, simplified river hydraulics and water quality are still used to predict the time of travel for pollutants along the reach, and the depth and velocity of flow used to predict re-aeration rates.

- Level 3

In level 3 studies calibrated flow routing models are used to more accurately predict time of travel along longer and more complex water bodies. This allows better representation of advective pollutant transport. More complex water quality simulation can be used with the model calibrated for the key parameters – BOD, ammonia and dissolved oxygen – using observed event sampling and water quality sonde data. Storm sewage quality is represented using observed sampling data or calibrated sewer quality models.

- Level 4

This is the most complex form of impact model. Calibrated hydrodynamic river models are used to simulate the varying depth and velocity of flow within the watercourse. Advection and dispersion is calibrated against observed data (e.g. dye tracing). Various levels of water quality simulation are possible with calibration and verification against event sampling and water quality sonde data.

For all levels, a long (minimum 10 year) historic or synthetic rainfall time series representative of the catchment is required.

New models are not required in all cases. Where they are 'fit for purpose', existing sewer and river impact models from recent drainage planning or UPM studies should be used.

Like the invertebrate quality assessment, the modelled water quality impact assessment uses a 'worst of' approach to classifying impact. Impact is assessed against both the 99 percentile and fundamental intermittent standards, and the worst impact classification is selected for input to the cost – benefit assessment. The FIS and 99 percentile standards provide alternate approaches to assessing the impact of storm discharges. The percentile standards seek to limit organic loading and total ammonia concentrations, while the FIS set acceptable limits for the determinands – dissolved oxygen and un-ionised ammonia – that are directly relevant to aquatic life. The levels of environmental protection the different standards set can't be compared directly (FWR, 2012). Both standards are subject to uncertainty in modelling, but the FIS are particularly vulnerable given the need to model a variety of additional water quality processes (UKTAG, 2012). Historically, both standards have been used as modelled water quality design standards to investigate and improve unsatisfactory overflows in response to observed water quality and / or ecological failures. The approach in the SOAF is slightly different to the traditional UPM approach. Rather than simply assessing whether the system is predicted to comply with water quality standards via a 'pass or fail' approach, which also considers confidence through sensitivity testing, the SOAF aims to quantify a range of impacts depending on the degree to which the FIS and 99 percentiles are exceeded (or not exceeded). This allows a wide range of impacts to be used in the benefits assessment. Both standards are assessed in a 'worst of approach', given that the comparable levels of protection afforded by the standards cannot be directly compared, and the levels of uncertainty involved in FIS modelling.

A) 99 percentile quality

To score impact against 99 percentile quality standards, two approaches are available depending on the type of modelling tool used:

1) Estimate of 99 percentile

The relevant 99 percentile BOD and total ammonia standards for the receiving water are selected according to WFD water body typology (FWR, 2012). Where modelling predicts the high frequency discharge to cause a drop in 99 percentile status class, an impact score of 45 is assigned. Where the overflow does not cause a drop in status class but causes a degree of within class deterioration, a score is assigned according to the percentage within class deterioration as shown in Table 11 below. The worst score returned for the BOD and total ammonia assessments is used to classify impact.

Table 10. Guidance on modelling approaches and levels of complexity.

Level	Urban drainage inputs				Boundary river conditions		River model		Rainfall series
	Storm sewage flow	**Storm sewage quality**	**WWTW flow**	**WWTW quality**	**Upstream river flow**	**Upstream river quality**	**Hydraulic**	**Water quality**	
1	Verified sewer model	Event mean concentrations using default values (e.g. Dempsey, 2005) or sampled values	Statistical distribution from MCertified data	Statistical distribution from sampled effluent quality	Statistical distribution from gauged data or ungauged estimate	Statistical distribution from EA routine samples	Simplified channel, steady & uniform	Simplified WQ processes & re-aeration using default values for rate coefficients	10 year representative historic or synthetic time series
2	Verified sewer model	Event mean concentrations using default values (e.g. Dempsey, 2005) or sampled values	Predicted flow time series from verified sewer model	Statistical distribution from sampled effluent quality	10 year historic flow time series from EA gauging station or calibrated rainfall runoff model	Statistical distribution from EA routine samples	Simplified channel, steady & uniform	Simplified WQ processes & re-aeration using default values for rate coefficients	10 year representative historic or synthetic time series
3	Verified sewer model	Sampled values or calibrated sewer quality model	Predicted flow time series from verified sewer model	Statistical distribution from sampled effluent quality	10 year historic flow time series from EA gauging station or calibrated rainfall runoff model	Statistical distribution from EA routine samples	Calibrated flow routing model	Advective pollutant transport, WQ simulation calibrated from event sampling & sonde data	10 year representative historic or synthetic time series
4	Verified sewer model	Sampled values or calibrated sewer quality model	Predicted flow time series from verified model	Statistical distribution from sampled effluent quality	10 year historic flow time series from EA gauging station or calibrated rainfall runoff model	Statistical distribution from EA routine samples	Calibrated hydrodynamic model	Calibrated advection – dispersion model, WQ simulation calibrated from event sampling & sonde data	10 year representative historic or synthetic time series

Table 11. 99th percentile within class deterioration scores.

Percentage within class deterioration	Score
1 – 10	5
11 – 25	15
26 – 50	25
51 – 75	35
>75	45

2) Duration of exceedance

Where simplified modelling tools are used which do not calculate a 99 percentile, but instead estimate the duration for which a 99 percentile standard is exceeded, then the following scoring system will be used in conjunction with the 99 percentile BOD and total ammonia standards for good status (see Table 12). The impact duration with the worst score should be used.

Table 12. Scoring system for duration / number of 99 percentile exceedances.

Impact duration	Allowable exceedances (no./year)	Score
1 hour	87.6	+ 0.50 points for every 1.0/yr increase in exceedances
6 hours	14.6	+ 3.0 points for every 1.0/yr increase in exceedances
24 hours	3.65	+ 12.0 points for every 1.0/yr increase in exceedances

B) Fundamental intermittent standards

To quantify impact against the FIS, the relevant standards for the receiving water are chosen according to fishery type (sustainable cyprinid, sustainable salmonid, and salmonid spawning) (FWR, 2012). The frequency of predicted FIS exceedances in the receiving water, with and without the discharge from the high frequency spiller are compared. The difference in the number of FIS exceedances predicted is used to quantify impact using the scoring system in Table 13, where the discharge causes a deterioration (increase) in the frequency of allowable exceedances:

Table 13. Scoring system for increases in FIS exceedances for un-ionised ammonia and dissolved oxygen.

Frequency (return period)	Allowable exceedances (no./year)	Score
1 month	12	+ 1.5 point for every 0.5/yr increase in exceedances
3 months	4	+ 4.0 points for every 0.5/yr increase in exceedances
1 year	1	+ 6.0 points for every 0.2/yr increase in exceedances

The worst score obtained from the FIS and 99 percentile assessments described above is then used to classify water quality impact as shown in Table 14 below. This impact classification will be used to input to the cost – benefit assessment.

Table 14. Water quality impact classification.

Water quality Score	Water quality impact classification
0 – 5	No impact
6 – 9	Very low
10 – 19	Low
20 – 29	Moderate
30 – 39	High
40 or more	Severe

4.0 Summary

The environmental impact assessment described above sets out a consistent process for quantifying the impact of a high frequency spiller. The use of a scoring system for classifying impact for aesthetics and invertebrates or water quality will provide information to inform an assessment of the costs and benefits of potential improvement options. By weighing the cost of technical solutions against the monetized benefits a more effective drainage system would provide, including wider socio – economic benefits as well as environmental impact, the SOAF should demonstrate compliance with Urban Waste requirements. In addition to compliance with relevant legislation, and as the UK leaves the EU, the SOAF will also demonstrate to the public and other stakeholders that the industry is proactively monitoring and investigating the performance of its storm overflows.

The environmental impact assessment described is still at a draft stage. Progress between July 2015 and autumn 2017 has been slow. Very little testing of the impact assessment has been carried out to date, and all companies agree much more testing is needed. Additional work also needs to consider how impact will be quantified for discharges to still freshwaters (i.e. lakes and canals), and coastal and transitional waters which are not designated bathing or shellfish waters.

References

Armitage, P. D., Moss, D., Wright, J. F., Furse, M. T. 1983. The performance of a new water quality score system based on macroinvertebrates over a wide range of unpolluted running water sites. Water Research, Vol. 17, pp. 333 – 347.

Benyon, Richard. 2013. Letter to Water and Sewerage Company Chief Executive Officers, 18 July 2013.

Chandler, J. R. 1970. A biological approach to water quality management. Water Pollution Control, Vol. 69, pp. 415 – 422.

Clarke, R. T., Wright, J. F., & Furse, M. T. 2003. RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. Ecological Modelling, Vol. 160, pp. 219 – 233.

Davy – Bowker, J., Clarke, R., Corbin, T., Vincent, H., Pretty, J., Hawczak, J., Blackburn, J., & Murphy, J. 2007. River Invertebrate Classification Tool. SNIFFER project WFD72c.

Davy – Bowker, J., Arnott, S., Close, R., Dobson, M., Dunbar, M., Jofre, G., Morton, D., Murphy, J., Wareham, W., Smith, S., & Gordon, V. 2010. Further development of River Classification Tool. Final Report, SNIFFER project WFD100.

Environment Agency. 2013. Risk based approach to the monitoring of storm discharges, 30 September 2013.

EU-STAR. 2004. UK invertebrate sampling and analysis procedure for STAR project, RIVPACS macroinvertebrate sampling protocol. Available at: http://www.eu-star.at/pdf/RivpacsMacroinvertebrateSamplingProtocol.pdf.

EUR-Lex. 2012. Judgment of the Court (First Chamber) of 18 October 2012. [Online] Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1508426296417&uri=CELEX:62010CJ0301.

EUR-Lex. 2017. Judgment of the Court (Ninth Chamber) of 4 May 2017. [Online] Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62015CJ0502.

Extence, C. A. & Ferguson, A. J. D. 1989. Aquatic invertebrate surveys as a water quality management tool in the Anglian water region. Regulated Rivers: Research and Management, Vol. 4, pp. 139 – 146.

FWR, 1994. Report No. FR0466 – User guide for assessing the impacts of combined sewer overflows. Foundation for Water Research, 1994.

FWR, 2012. Urban Pollution Management Manual – Third Edition. Foundation for Water Research.

Marine Conservation Society, 2011. Combined sewer overflows pollution policy and position statement. Version 1, October 2011. [Online] Available at: https://www.mcsuk.org/downloads/pollution/CSO%20policy.pdf.

Paisley, M. F., Trigg, D.J., & Whalley, W. J. 2007. Revision and testing of BMWP scores. Final report SNIFFER project WFD72a. Edinburgh, SNIFFER.

UKCCRA, 2017. United Kingdom Climate Change Risk Assessment 2017 Evidence Report. [Online] Available at: https://www.theccc.org.uk/tackling-climate-change/preparing-for-climate-change/uk-climate-change-risk-assessment-2017.

UKTAG, 2012. Updated recommendations on environmental standards: river basin management (2015 – 2021). Water Framework Directive – United Kingdom Technical Advisory Group (WFD – UKTAG), April 2012.

UKTAG, 2014. UKTAG River assessment method benthic invertebrate fauna: invertebrates (general degradation): Whalley, Hawkes, Paisley & Trigg (WHPT) metric in River Invertebrate Classification Tool (RICT). Water Framework Directive – United Kingdom Technical Advisory Group (WFD – UKTAG), July 2014.

Water UK, 2017. 21st Century Drainage Programme. [Online] https:/www.water.org.uk/policy/improving-resilience/21st-century-drainage.

Whalley, W. J., & Hawkes, H. A. 1997. A computer – based development of the biological working party monitoring score system incorporating abundance rating, site type and indicator value. Water Research, Vol. 31. No.2. pp201 – 210.

Woodiwiss, F. S. 1964. The biological system of stream classification used by the River Trent Board. Chemistry Industry, Vol. 11, pp. 443 – 447.

Wright, J. F., Sutcliffe, D. W., & Furse, M. T. (Eds). 2000. Assessing the biological quality of freshwaters: RIVPACS and other techniques. Freshwater Biological Association, Ambleside.